

# Recovery of Block-Sparse Representations from Noisy Observations via Orthogonal Matching Pursuit

Jun Fang, *Member, IEEE*, and Hongbin Li, *Senior Member, IEEE*

**Abstract**— We study the problem of recovering the sparsity pattern of block-sparse signals from noise-corrupted measurements. A simple, efficient recovery method, namely, a block-version of the orthogonal matching pursuit (OMP) method, is considered in this paper and its behavior for recovering the block-sparsity pattern is analyzed. We provide sufficient conditions under which the block-version of the OMP can successfully recover the block-sparse representations in the presence of noise. Our analysis reveals that exploiting block-sparsity can improve the recovery ability and lead to a guaranteed recovery for a higher sparsity level. Numerical results are presented to corroborate our theoretical claim.

**Index Terms**—Block-sparsity, orthogonal matching pursuit, compressed sensing.

## I. INTRODUCTION

The problem of recovering a high dimensional sparse signal based on a small number of measurements has been of significant interest in signal and imaging processing, applied mathematics, and statistics. Such a problem arises from a number of applications, including subset selection in regression [1], structure estimation in graphical models [2], and compressed sensing [3]. Among these applications, many involves determining the locations of the nonzero components of the sparse signal, which is also referred to as sparsity pattern recovery (or more simply, sparsity recovery). In practice, the locations of the nonzero components (or, the support of the sparse signals) usually have significant physical meanings. For example, in chemical agent detection, the indices for the nonzero coordinates indicates the chemical components present in a mixture. In sparse linear regression, the recovered support corresponds to a small subset of features which linearly influence the observed data. Due to its importance, sparsity pattern recovery has received considerable attention over the past few years. In [4], [5], the authors analyzed the behavior of  $\ell_1$ -constrained quadratic programming (QP), also referred to as the Lasso, for recovering the sparsity pattern in a deterministic framework. Sufficient conditions were established for exact sparsity pattern recovery. Such a problem was also studied in [6] from a statistical perspective, where necessary and sufficient conditions on the problem dimension, the number of nonzero elements, and the number

of measurements are established for sparsity pattern recovery. Recently, information-theoretic limits of sparsity recovery with an exhaustive search decoder were studied in [7], [8].

In this paper, we consider the problem of recovering block-sparse signals whose nonzero elements appear in fixed blocks. Block-sparse signals arise naturally. For example, the atomic decomposition of multi-band signals [9] or audio signals [10] usually results in a block-sparse structure in which the nonzero coefficients occur in clusters. Recovery of block-sparse signals has been extensively studied in [11]–[13], in which the recovery behaviors of the basis pursuit (BP), or  $\ell_1$ -constrained QP, and the orthogonal matching pursuit (OMP) algorithms were analyzed via the restricted isometry property (RIP) [12], [13] and the mutual coherence property [11]. Their analyses [11]–[13] revealed that exploiting block-sparsity yields a relaxed condition which can guarantee recovery for a higher sparsity level as compared with treating block-sparse signals as conventional sparse signals. Nevertheless, most of these studies focused on noiseless scenarios. In practice, measurements are inevitably contaminated with noise and underlying uncertainties. It is therefore important to analyze the effect of measurement noise on the block-sparsity pattern recovery, e.g. under what conditions the exact sparsity pattern can be recovered, and does exploiting block-sparsity still lead to a guaranteed recovery for a higher sparsity level? These questions will be addressed in this paper. Specifically, we consider a block version of the OMP algorithm and study its behavior for recovering block-sparsity pattern in the presence of noise. A comparison with the theoretical results for the conventional OMP algorithm [5] is presented to highlight the benefits of exploiting block-sparsity property.

## II. PROBLEM FORMULATION

We consider the problem of recovering a block-sparse signal  $\mathbf{x} \in \mathbb{R}^n$  from noise-corrupted measurements

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w} \quad (1)$$

where  $\mathbf{A} \in \mathbb{R}^{m \times n}$  ( $m < n$ ) is the measurement matrix with unit-norm columns, and  $\mathbf{w}$  is an arbitrary and unknown vector of errors. To define block-sparsity, as in [11], we model  $\mathbf{x}$  as a concatenation of equal-length blocks

$$\mathbf{x} = [\mathbf{x}_1^T \ \mathbf{x}_2^T \ \dots \ \mathbf{x}_L^T]^T \quad (2)$$

where  $\mathbf{x}_l \triangleq [x_{(l-1)d+1} \ \dots \ x_{ld}]^T$  is a  $d$ -dimensional vector. Clearly, the vector  $\mathbf{x}$  has a dimension  $n = Ld$ , and the

Jun Fang and Hongbin Li are with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ 07030, USA. E-mails: {Jun.Fang@ieee.org, Hongbin.Li@stevens.edu}

This work was supported in part by the National Science Foundation under Grant ECCS-0901066 and by the Air Force Office of Scientific Research under Grant FA9550-09-1-0310.

vector is called block  $K$ -sparse if its block component  $\mathbf{x}_l$  has nonzero Euclidean norm for at most  $K$  indices  $l$ . Similarly, the measurement matrix  $\mathbf{A}$  can be expressed as a concatenation of column-block matrices  $\{\mathbf{A}_l\}_{l=1}^L$

$$\mathbf{A} = [\mathbf{A}_1 \ \mathbf{A}_2 \ \dots \ \mathbf{A}_L] \quad (3)$$

where  $\mathbf{A}_l \in \mathbb{R}^{m \times d}$ . Also, we assume that the number of rows of  $\mathbf{A}$  is an integer multiples of  $d$ , i.e.  $m = Rd$  with  $R$  an integer. The conventional coherence metric of the measurement matrix  $\mathbf{A}$  is defined as

$$\mu \triangleq \max_{i \neq j} |\mathbf{a}_i^T \mathbf{a}_j| \quad (4)$$

where  $\mathbf{a}_i$  denotes the  $i$ th column of  $\mathbf{A}$ . This coherence metric, albeit useful, is not sufficient to characterize the block-structure of the sparse signal. To exploit the block-sparsity property, we define the block-coherence  $\mu_B$  and sub-coherence  $\nu$  (these two concepts were firstly introduced in [11]):

$$\begin{aligned} \mu_B &\triangleq \max_{i,j \neq i} \frac{1}{d} \rho(\mathbf{A}_i^T \mathbf{A}_j) \\ \nu &\triangleq \max_l \max_{i,j \neq i} |\mathbf{a}_i^T \mathbf{a}_j|, \quad \mathbf{a}_i, \mathbf{a}_j \in \mathbf{A}_l \end{aligned} \quad (5)$$

where  $\rho(\mathbf{X})$  denotes the spectral norm of  $\mathbf{X}$ , which is defined as the square root of the maximum eigenvalue of  $\mathbf{X}^T \mathbf{X}$ , i.e.  $\sqrt{\lambda_{\max}(\mathbf{X}^T \mathbf{X})}$ . Related properties of the block-coherence  $\mu_B$  can be found in [11]. We see that  $\mu_B$  quantifies the coherence between blocks of  $\mathbf{A}$ , while the coherence within blocks is characterized by the sub-coherence  $\nu$ .

The objective of this paper is to identify sufficient conditions on the measurement matrix  $\mathbf{A}$  (in terms of the block-coherence  $\mu_B$  and the sub-coherence  $\nu$ ), as well as the signal vector  $\mathbf{x}$  and the error vector  $\mathbf{w}$ , under which the block-sparsity pattern can be recovered from the noisy measurements. We are particularly interested in analyzing the recovery ability of a block-version of the orthogonal matching pursuit (OMP). OMP is a simple greedy approximation algorithm developed in [14], [15]. Despite its simplicity, OMP is a provably good approximation algorithm which achieves performance close to Lasso in certain scenarios [16], [17]. In the following, we briefly summarize the block-version of the OMP, which is also termed as block-OMP (BOMP). This BOMP is a slight variant of the original BOMP that was introduced in [11] for noiseless scenarios.

*BOMP Algorithm:*

- 1) Initialize the residual  $\mathbf{r}_0 = \mathbf{y}$ , the index set  $S_0 = \emptyset$ .
- 2) At the  $t$ th step ( $t \geq 1$ ), we choose the block that is best matched to  $\mathbf{r}_{t-1}$  according to

$$i_t = \arg \max_i \|\mathbf{A}_i^T \mathbf{r}_{t-1}\|_2 \quad (6)$$

- 3) Augment the index set and the matrix of chosen blocks:  $S_t = S_{t-1} \cup \{i_t\}$  and  $\Psi^{(t)} = [\Psi^{(t-1)} \ \mathbf{A}_{i_t}]$ . We use the convention that  $\Psi^{(0)}$  is an empty matrix.
- 4) Solve a least squares problem to obtain a new signal estimate  $\mathbf{x}_t = \arg \min_{\mathbf{x}} \|\mathbf{y} - \Psi^{(t)} \mathbf{x}\|_2$
- 5) Calculate the new residual as  $\mathbf{r}_t = \mathbf{y} - \Psi^{(t)} \mathbf{x}_t = \mathbf{y} - \mathcal{P}_{\Psi^{(t)}} \mathbf{y}$ , where  $\mathcal{P}_{\Psi^{(t)}} = \Psi^{(t)} (\Psi^{(t)})^\dagger$  is the orthogonal projection onto the column space of  $\Psi^{(t)}$ , and  $^\dagger$  stands for the pseudo-inverse.

- 6) If  $\|\mathbf{r}_t\|_2 \geq \epsilon$ , return to Step 2; otherwise stop.

### III. BLOCK-SPARSITY PATTERN RECOVERY ANALYSIS

Let  $\mathbf{x}_{\text{nz}}$  denote a  $Kd$  dimensional column vector constructed by stacking the nonzero block components  $\mathbf{x}_l, \forall \{l | \mathbf{x}_l \neq \mathbf{0}\}$ ,  $\mathbf{A}_{\text{nz}} \in \mathbb{R}^{m \times Kd}$  denote a submatrix of  $\mathbf{A}$  constructed by concatenating the column-blocks  $\mathbf{A}_l, \forall \{l | \mathbf{x}_l \neq \mathbf{0}\}$ , i.e. the blocks corresponding to the nonzero  $\mathbf{x}_l$ , and let  $\mathbf{A}_z \in \mathbb{R}^{m \times (L-K)d}$  stand for a submatrix of  $\mathbf{A}$  constructed by concatenating the column-blocks  $\mathbf{A}_l$  corresponding to zero  $\mathbf{x}_l$ . For notational convenience, let  $I_1 = \{l_1, l_2, \dots, l_K\}$  denote a set of indices for which  $\mathbf{x}_{l_i} \neq \mathbf{0}$ , and  $I_2 = \{l_{K+1}, l_{K+2}, \dots, l_L\}$  denote a set of indices for which  $\mathbf{x}_{l_i} = \mathbf{0}$ . Therefore we can write

$$\begin{aligned} \mathbf{x}_{\text{nz}} &\triangleq [\mathbf{x}_{l_1}^T \ \mathbf{x}_{l_2}^T \ \dots \ \mathbf{x}_{l_K}^T]^T \\ \mathbf{A}_{\text{nz}} &\triangleq [\mathbf{A}_{l_1} \ \mathbf{A}_{l_2} \ \dots \ \mathbf{A}_{l_K}] \\ \mathbf{A}_z &\triangleq [\mathbf{A}_{l_{K+1}} \ \mathbf{A}_{l_{K+2}} \ \dots \ \mathbf{A}_{l_L}] \end{aligned}$$

The measurements can therefore be written as

$$\mathbf{y} = \mathbf{A}_{\text{nz}} \mathbf{x}_{\text{nz}} + \mathbf{w} \quad (7)$$

We can decompose the error vector  $\mathbf{w}$  into  $\mathbf{w} = \mathcal{P}_{\mathbf{A}_{\text{nz}}} \mathbf{w} + \mathcal{P}_{\mathbf{A}_{\text{nz}}}^\perp \mathbf{w}$ , where  $\mathcal{P}_{\mathbf{A}_{\text{nz}}} = \mathbf{A}_{\text{nz}} \mathbf{A}_{\text{nz}}^\dagger$  denotes the orthogonal projection onto the subspace spanned by the columns of  $\mathbf{A}_{\text{nz}}$ , and  $\mathcal{P}_{\mathbf{A}_{\text{nz}}}^\perp = \mathbf{I} - \mathcal{P}_{\mathbf{A}_{\text{nz}}}$  is the orthogonal projection onto the null space of  $\mathbf{A}_{\text{nz}}^T$ . We can further write

$$\begin{aligned} \mathbf{y} &= \mathbf{A}_{\text{nz}} \mathbf{x}_{\text{nz}} + \mathbf{w} = \mathbf{A}_{\text{nz}} \mathbf{x}_{\text{nz}} + \mathcal{P}_{\mathbf{A}_{\text{nz}}} \mathbf{w} + \mathcal{P}_{\mathbf{A}_{\text{nz}}}^\perp \mathbf{w} \\ &= \mathbf{A}_{\text{nz}} (\mathbf{x}_{\text{nz}} + \mathbf{A}_{\text{nz}}^\dagger \mathbf{w}) + \mathcal{P}_{\mathbf{A}_{\text{nz}}}^\perp \mathbf{w} \\ &\triangleq \mathbf{A}_{\text{nz}} \tilde{\mathbf{x}}_{\text{nz}} + \tilde{\mathbf{w}} \end{aligned} \quad (8)$$

where  $\tilde{\mathbf{x}}_{\text{nz}} \triangleq \mathbf{x}_{\text{nz}} + \mathbf{A}_{\text{nz}}^\dagger \mathbf{w}$ , and  $\tilde{\mathbf{w}} \triangleq \mathcal{P}_{\mathbf{A}_{\text{nz}}}^\perp \mathbf{w}$ . Equation (8) decomposes the measurements into two mutually orthogonal components: a signal component  $\mathbf{A}_{\text{nz}} \tilde{\mathbf{x}}_{\text{nz}}$  and a noise component  $\tilde{\mathbf{w}}$ . The reason for doing so is that even the exact signal support (block-sparsity pattern) is known, there is no way to separate the noise projection term  $\mathbf{A}_{\text{nz}}^\dagger \mathbf{w}$  from the true signal  $\mathbf{x}_{\text{nz}}$ . Hence it is more convenient to carry out our analysis based on (8) instead of (7).

Recall that, at each iteration, the BOMP algorithm searches for a block that is best matched to the residual vector according to (6). We can define a greedy selection ratio that determines whether or not a correct block is selected at each iteration

$$\gamma_t = \frac{\max_{l \in I_2} \|\mathbf{A}_l^T \mathbf{r}_{t-1}\|_2}{\max_{l \in I_1} \|\mathbf{A}_l^T \mathbf{r}_{t-1}\|_2} \quad (9)$$

where  $\mathbf{r}_{t-1}$  is the residual vector at iteration  $t-1$ . Clearly, at each iteration, the algorithm picks an index whose corresponding block is in  $\mathbf{A}_{\text{nz}}$  if  $\gamma_t < 1$ , otherwise an incorrect index whose corresponding block is in  $\mathbf{A}_z$  is chosen. Since the residual is orthogonal to the subspace spanned by all the previously chosen block-columns, no index will be chosen twice. Therefore, in order to recover the block-sparsity pattern, we need to guarantee  $\gamma_t < 1$  throughout the first  $K$  iterations, i.e.  $\gamma_t < 1, \forall t \leq K$ . Here for simplicity, we assume that the number of nonzero blocks,  $K$ , is known *a priori*. In practice,  $K$  can be automatically determined by the BOMP algorithm given the error tolerance  $\epsilon$  ( $\epsilon$  can be estimated from

the observation noise power in practice). As long as  $K$  is not overestimated, i.e.  $\hat{K} \leq K$ , we can ensure that all the chosen indices are from the set of correct indices  $I_1$ .

In the following, we derive sufficient conditions that guarantee  $\gamma_t < 1$  throughout the first  $K$  iterations. Before proceeding, we define a general mixed  $\ell_2/\ell_p$ -norm ( $p = 1, 2, \infty$ ) that will be used throughout this paper. For a vector  $\mathbf{z} = [\mathbf{z}_1^T \ \mathbf{z}_2^T \ \dots \ \mathbf{z}_Q^T]^T$  consisting of equal-length blocks with block size  $d$ , the general mixed  $\ell_2/\ell_p$ -norm (with block size  $d$ ) is defined as

$$\|\mathbf{z}\|_{2,p} = \|\mathbf{v}\|_p \quad \text{where } v_q = \|\mathbf{z}_q\|_2 \quad (10)$$

Correspondingly, for a matrix  $\mathbf{X} \in \mathbb{R}^{U \times Qd}$ , where  $U$  and  $Q$  can be any positive integers, the mixed matrix norm (with block size  $d$ ) is defined as

$$\|\mathbf{X}\|_{2,p} = \max_{\mathbf{z} \neq \mathbf{0}} \frac{\|\mathbf{X}\mathbf{z}\|_{2,p}}{\|\mathbf{z}\|_{2,p}} \quad (11)$$

Resorting to this general mixed  $\ell_2/\ell_p$ -norm (with block size  $d$ ) definition, the greedy selection ratio defined in (9) can be re-expressed as

$$\gamma_t = \frac{\max_{\{l: \mathbf{x}_l = \mathbf{0}\}} \|\mathbf{A}_l^T \mathbf{r}_{t-1}\|_2}{\max_{\{l: \mathbf{x}_l \neq \mathbf{0}\}} \|\mathbf{A}_l^T \mathbf{r}_{t-1}\|_2} = \frac{\|\mathbf{A}_z^T \mathbf{r}_{t-1}\|_{2,\infty}}{\|\mathbf{A}_{nz}^T \mathbf{r}_{t-1}\|_{2,\infty}} \quad (12)$$

Suppose that the BOMP algorithm has successfully executed the first  $k$  ( $k < K$ ) iterations with residual

$$\mathbf{r}_k = \mathbf{y} - \mathcal{P}_{\Phi_1} \mathbf{y} \quad (13)$$

where  $\Phi_1 \in \mathbb{R}^{m \times kd}$  is a matrix constructed by concatenating the  $k$  block-columns chosen from the previous  $k$  iterations, and  $\mathcal{P}_{\Phi_1} = \Phi_1 \Phi_1^\dagger$  is the orthogonal projection onto the column space of  $\Phi_1$ . Note that  $\Phi_1$  is a sub-matrix of  $\mathbf{A}_{nz}$  since we assume that the algorithm selected the correct indices during the first  $k$  iterations. Let  $\Phi_2$  be a matrix constructed by concatenating the remaining  $K - k$  column-blocks in  $\mathbf{A}_{nz}$ . Without loss of generality, we can write  $\mathbf{A}_{nz} = [\Phi_1 \ \Phi_2]$ , i.e.  $\Phi_1 \triangleq [\mathbf{A}_{l_1} \ \dots \ \mathbf{A}_{l_k}]$ , and  $\Phi_2 \triangleq [\mathbf{A}_{l_{k+1}} \ \dots \ \mathbf{A}_{l_K}]$ . Also, we write  $\tilde{\mathbf{x}}_{nz} = [\tilde{\mathbf{x}}_{l_1}^T \ \tilde{\mathbf{x}}_{l_1}^T \ \dots \ \tilde{\mathbf{x}}_{l_K}^T]^T = [\phi_1^T \ \phi_2^T]^T$ , where  $\phi_1 \triangleq [\tilde{\mathbf{x}}_{l_1}^T \ \dots \ \tilde{\mathbf{x}}_{l_k}^T]^T$ , and  $\phi_2 \triangleq [\tilde{\mathbf{x}}_{l_{k+1}}^T \ \dots \ \tilde{\mathbf{x}}_{l_K}^T]^T$ . Substituting (8) into (13), the residual can be written as

$$\begin{aligned} \mathbf{r}_k &= \mathbf{A}_{nz} \tilde{\mathbf{x}}_{nz} + \tilde{\mathbf{w}} - \mathcal{P}_{\Phi_1} (\mathbf{A}_{nz} \tilde{\mathbf{x}}_{nz} + \tilde{\mathbf{w}}) \\ &\stackrel{(a)}{=} \mathbf{A}_{nz} \tilde{\mathbf{x}}_{nz} - \mathcal{P}_{\Phi_1} \mathbf{A}_{nz} \tilde{\mathbf{x}}_{nz} + \tilde{\mathbf{w}} \\ &\stackrel{(b)}{=} \Phi_2 \phi_2 - \mathcal{P}_{\Phi_1} \Phi_2 \phi_2 + \tilde{\mathbf{w}} \\ &\stackrel{(c)}{=} \tilde{\mathbf{r}}_k + \tilde{\mathbf{w}} \end{aligned} \quad (14)$$

where (a) comes from the fact that  $\tilde{\mathbf{w}}$  is orthogonal to the column space of  $\Phi_1$ , and (b) comes by noting that  $\mathcal{P}_{\Phi_1} \Phi_1 = \Phi_1$ , and in (c) we define  $\tilde{\mathbf{r}}_k \triangleq \Phi_2 \phi_2 - \mathcal{P}_{\Phi_1} \Phi_2 \phi_2$ . Using this

result, the greedy selection ratio at iteration  $k + 1$  becomes

$$\begin{aligned} \gamma_{k+1} &= \frac{\|\mathbf{A}_z^T \mathbf{r}_k\|_{2,\infty}}{\|\mathbf{A}_{nz}^T \mathbf{r}_k\|_{2,\infty}} = \frac{\|\mathbf{A}_z^T (\tilde{\mathbf{r}}_k + \tilde{\mathbf{w}})\|_{2,\infty}}{\|\mathbf{A}_{nz}^T (\tilde{\mathbf{r}}_k + \tilde{\mathbf{w}})\|_{2,\infty}} \\ &= \frac{\|\mathbf{A}_z^T (\tilde{\mathbf{r}}_k + \tilde{\mathbf{w}})\|_{2,\infty}}{\|\mathbf{A}_{nz}^T \tilde{\mathbf{r}}_k\|_{2,\infty}} \stackrel{(a)}{\leq} \frac{\|\mathbf{A}_z^T \tilde{\mathbf{r}}_k\|_{2,\infty} + \|\mathbf{A}_z^T \tilde{\mathbf{w}}\|_{2,\infty}}{\|\mathbf{A}_{nz}^T \tilde{\mathbf{r}}_k\|_{2,\infty}} \\ &\stackrel{(b)}{=} \frac{\|\mathbf{A}_z^T \mathcal{P}_{\mathbf{A}_{nz}} \tilde{\mathbf{r}}_k\|_{2,\infty}}{\|\mathbf{A}_{nz}^T \tilde{\mathbf{r}}_k\|_{2,\infty}} + \frac{\|\mathbf{A}_z^T \tilde{\mathbf{w}}\|_{2,\infty}}{\|\mathbf{A}_{nz}^T \tilde{\mathbf{r}}_k\|_{2,\infty}} \\ &= \frac{\|\mathbf{A}_z^T (\mathbf{A}_{nz}^\dagger)^T \mathbf{A}_{nz}^T \tilde{\mathbf{r}}_k\|_{2,\infty}}{\|\mathbf{A}_{nz}^T \tilde{\mathbf{r}}_k\|_{2,\infty}} + \frac{\|\mathbf{A}_z^T \tilde{\mathbf{w}}\|_{2,\infty}}{\|\mathbf{A}_{nz}^T \tilde{\mathbf{r}}_k\|_{2,\infty}} \\ &\leq \|\mathbf{A}_z^T (\mathbf{A}_{nz}^\dagger)^T\|_{2,\infty} + \frac{\|\mathbf{A}_z^T \tilde{\mathbf{w}}\|_{2,\infty}}{\|\mathbf{A}_{nz}^T \tilde{\mathbf{r}}_k\|_{2,\infty}} \end{aligned} \quad (15)$$

where (a) comes from the fact the general mixed  $\ell_2/\ell_p$ -norm satisfies the triangle inequality:  $\|\mathbf{a} + \mathbf{b}\|_{2,\infty} \leq \|\mathbf{a}\|_{2,\infty} + \|\mathbf{b}\|_{2,\infty}$ , which can be readily verified, (b) follows from  $\mathcal{P}_{\mathbf{A}_{nz}} \tilde{\mathbf{r}}_k = \tilde{\mathbf{r}}_k$  since  $\tilde{\mathbf{r}}_k$  lies in the column space of  $\mathbf{A}_{nz}$ . Our objective is to identify conditions assuring  $\gamma_{k+1} < 1$ .

If the measurement process is perfect and noise-free, that is,  $\mathbf{y} = \mathbf{A}\mathbf{x}$ , then the greedy selection ratio is simply upper bounded by

$$\gamma_{k+1} \leq \|\mathbf{A}_z^T (\mathbf{A}_{nz}^\dagger)^T\|_{2,\infty} \quad (16)$$

Furthermore, it has been shown in [11, Lemma 4] that  $\|\mathbf{A}_z^T (\mathbf{A}_{nz}^\dagger)^T\|_{2,\infty}$  is upper bounded by

$$\|\mathbf{A}_z^T (\mathbf{A}_{nz}^\dagger)^T\|_{2,\infty} \leq \frac{K d \mu_B}{1 - (d-1)\nu - (K-1)d\mu_B} \quad (17)$$

Therefore the condition  $\gamma_{k+1} < 1$  holds universally if the block-coherence  $\mu_B$  and sub-coherence  $\nu$  associated with the dictionary  $\mathbf{A}$  satisfies

$$\frac{K d \mu_B}{1 - (d-1)\nu - (K-1)d\mu_B} < 1 \quad (18)$$

Since, in practice, measurements are inevitably contaminated with noise and underlying uncertainties, it is thus important to understand the effect of measurement noise on the block-sparsity pattern recovery. Apparently, when noise is present, condition (18) alone cannot guarantee the exact recovery of the block-sparsity pattern. Instead, from (15), we see that, to assure  $\gamma_{p+1} < 1$ , we need

$$\|\mathbf{A}_z^T (\mathbf{A}_{nz}^\dagger)^T\|_{2,\infty} + \frac{\|\mathbf{A}_z^T \tilde{\mathbf{w}}\|_{2,\infty}}{\|\mathbf{A}_{nz}^T \tilde{\mathbf{r}}_k\|_{2,\infty}} < 1 \quad (19)$$

The inequality (19) has to hold valid for  $0 \leq k \leq K - 1$  in order to ensure that the BOMP algorithm chooses the correct indices throughout the first  $K$  iterations. In the following, we provide sufficient conditions which guarantee (19) for  $0 \leq k \leq K - 1$ . The results are summarized as follows.

*Theorem 1:* Let

$$\omega \triangleq \|\mathbf{A}^T \tilde{\mathbf{w}}\|_{2,\infty} = \max_l \|\mathbf{A}_l^T \tilde{\mathbf{w}}\|_2 \quad (20)$$

denote the maximum correlation between the column block  $\mathbf{A}_l$  and the residual noise component  $\tilde{\mathbf{w}}$ . Let

$$x_{b,\min} \triangleq \min_{l \in I_1} \|\tilde{\mathbf{x}}_l\|_2 \quad (21)$$

the minimum  $\ell_2$ -norm of the non-zero signal block components. Suppose that the following conditions are satisfied

$$\begin{aligned} \text{(i)} \quad & 1 - (d-1)\nu - (2K-1)d\mu_B > 0 \\ \text{(ii)} \quad & \frac{[1 - (d-1)\nu - (2K-1)d\mu_B]^2}{1 - (d-1)\nu - (K-1)d\mu_B} > \frac{\omega}{x_{b,\min}} \end{aligned} \quad (22)$$

then we can guarantee that the BOMP algorithm selects indices from  $I_1$  throughout the first  $K$  iterations. If the error tolerance  $\epsilon$  is chosen such that the algorithm stops at the end of iteration  $K$ , then the BOMP recovers the exact block-sparsity pattern.

*Proof:* See Appendix A. ■

Theorem 1 is a generalization of the results presented in [11] which considered block-sparse signal recovery from noise-free measurements. To see this, for the noiseless case, we have  $\omega = 0$ , and hence the condition (22) is simplified as

$$1 - (d-1)\nu - (2K-1)d\mu_B > 0 \quad (23)$$

which is exactly the recovery condition provided in [11] for block-sparse signal recovery. On the other hand, for the noisy case, the success of the BOMP algorithm not only depends on the block-coherence  $\mu_B$  and the sub-coherence  $\nu$ , but also depends on the ratio of the maximum correlation (between the column block  $\mathbf{A}_l$  and the residual noise component  $\tilde{\mathbf{w}}$ ) to the minimum  $\ell_2$ -norm of the nonzero signal block components  $\tilde{\mathbf{x}}_l, \forall l \in I_1$ . The importance of the minimum nonzero signal component in sparsity pattern recovery has been highlighted in [7], [8]. In particular, [7] showed that both the sufficient and necessary conditions require control of the minimum nonzero signal component. Our result suggests that, for block-sparse signal recovery, the minimum  $\ell_2$ -norm of the nonzero signal block components, instead of the minimum magnitude of an entry, is the key quantity that controls the block subset selection.

Also, we observe that the left-hand side of the second condition in (22) is strictly less than one. Therefore the ratio  $\omega/x_{b,\min}$  cannot be greater than one, otherwise the condition cannot be met, irrespective of the choice of the sub-coherence  $\nu$  and the block-coherence  $\mu_B$ . The deterministic condition (22), however, guarantees recovery of the sparsity pattern under the worst-case scenario and therefore is very pessimistic. If we take a probabilistic analysis (as in [18]) that ensures a probabilistic recovery, the condition can be significantly relaxed. This could be a direction of our future study.

#### IV. DISCUSSIONS

We note that in this paper, as in [11], block-sparsity is explicitly exploited to yield a more relaxed condition imposed on the measurement matrix, and therefore lead to a guaranteed recovery for a potentially higher sparsity level. If the block-sparse signal is treated as a conventional  $Kd$ -sparse vector without exploiting knowledge of the block-sparsity structure, sufficient conditions for exact sparsity pattern recovery using OMP are given in [5, Theorem 18] and can be formulated as (by combining the first and the third equation in [5, Theorem

18])

$$\begin{aligned} \text{(i)} \quad & 1 - 2Kd\mu > 0 \\ \text{(ii)} \quad & \frac{(1 - 2Kd\mu)^2}{1 - Kd\mu} > \frac{\|\mathbf{A}^T \tilde{\mathbf{w}}\|_\infty}{x_{\min}} \end{aligned} \quad (24)$$

where  $x_{\min}$  denotes the minimum magnitude of the nonzero signal elements in  $\tilde{\mathbf{x}}_{nz}$ . When  $d = 1$ , block-sparsity reduces to conventional sparsity and we have  $\nu = 0$ ,  $\mu_B = \mu$ . The condition (22) is simplified as

$$\begin{aligned} \text{(i)} \quad & 1 - (2K-1)d\mu > 0 \\ \text{(ii)} \quad & \frac{(1 - (2K-1)d\mu)^2}{1 - (K-1)d\mu} > \frac{\|\mathbf{A}^T \tilde{\mathbf{w}}\|_\infty}{x_{\min}} \end{aligned} \quad (25)$$

which is the same as (24) except that  $2K$  and  $K$  in the numerator and denominator are replaced by  $2K-1$  and  $K-1$ , respectively (It can be easily verified that (25) is slightly loose than (24)). When  $d > 1$ , in the special case that the columns of  $\mathbf{A}_l$  are orthonormal for each  $l$ , we have  $\nu = 0$  and therefore the recovery condition (22) becomes

$$\begin{aligned} \text{(i)} \quad & 1 - (2K-1)d\mu_B > 0 \\ \text{(ii)} \quad & \frac{[1 - (2K-1)d\mu_B]^2}{1 - (K-1)d\mu_B} > \frac{\omega}{s_{\min}} \end{aligned} \quad (26)$$

This recovery condition, (26), is less restrictive than (24) since we have

$$\begin{aligned} \frac{[1 - (2K-1)d\mu_B]^2}{1 - (K-1)d\mu_B} &> \frac{(1 - 2Kd\mu_B)^2}{1 - Kd\mu_B} \stackrel{(a)}{\geq} \frac{(1 - 2Kd\mu)^2}{1 - Kd\mu} \\ &> \frac{\|\mathbf{A}^T \tilde{\mathbf{w}}\|_\infty}{x_{\min}} \stackrel{(b)}{\geq} \frac{\omega}{x_{b,\min}} \end{aligned} \quad (27)$$

where (a) comes from the fact that  $1 - 2Kd\mu > 0$  and  $\mu_B \leq \mu$  [11, Proposition 2], (b) follows from  $\omega \leq \sqrt{d}\|\mathbf{A}^T \tilde{\mathbf{w}}\|_\infty$  and  $x_{b,\min} \geq \sqrt{d}x_{\min}$ . We see that through exploiting the block-sparsity, the sparsity pattern recovery condition is relaxed and we can guarantee a recovery of sparsity pattern with a higher sparsity level. A close examination of (27) reveals that this improvement comes from two aspects. First, the measurement matrix requires a less restrictive mutual coherence condition since  $\mu_B \leq \mu$ . Second, for the same signal, noise, and measurement matrix, the quantity  $\omega/x_{b,\min}$  is always smaller than or equal to  $\|\mathbf{A}^T \tilde{\mathbf{w}}\|_\infty/x_{\min}$ , meaning that exploiting block-sparsity can improve the ability of detecting weak signals buried in noise.

If the individual blocks  $\mathbf{A}_l$  are, however, not orthonormal, then  $\nu > 0$ , and  $\nu$  has to be small in order to result in a performance gain for block-sparsity recovery as compared with the conventional sparse recovery. We can also follow the orthogonalization approach [11] to analyze the general non-orthonormal case. We orthogonalize the individual blocks  $\mathbf{A}_l = \tilde{\mathbf{A}}_l \mathbf{V}_l$ , in which  $\tilde{\mathbf{A}}_l$  consists of orthonormal columns, and  $\mathbf{V}_l$  is an invertible matrix. The original dictionary can therefore be written as  $\mathbf{A} = \tilde{\mathbf{A}} \mathbf{V}$ , where  $\mathbf{V}$  is a block-diagonal matrix with blocks  $\mathbf{V}_l$ . Clearly, orthogonalization preserves the block-sparsity level. The comparison that is meaningful here is between the recovery based on the original model without exploiting block-sparsity and the recovery based on the orthogonalized model taking block-sparsity into account.

For the orthogonalized dictionary  $\tilde{\mathbf{A}}$ , we have  $\nu(\tilde{\mathbf{A}}) = 0$ . Therefore we are only concerned about the relation between  $\mu$  before orthogonalization and  $\mu_B$  after orthogonalization, which are denoted by  $\mu(\mathbf{A})$  and  $\mu_B(\tilde{\mathbf{A}})$  respectively. Although an exact relation between  $\mu(\mathbf{A})$  and  $\mu_B(\tilde{\mathbf{A}})$  is difficult to derive, it has been shown in [11] that if  $d > RL/(L - R)$ , then we have  $\mu(\mathbf{A}) \geq \mu_B(\tilde{\mathbf{A}})$ . Hence even for general dictionaries, exploiting block-sparsity still leads to a guaranteed sparsity pattern recovery for a potentially higher sparsity level by properly choosing the number of measurements to satisfy  $d > RL/(L - R)$ .

We explore the connection and difference between our work and [19], [20]. In [19], [20], the problem of simultaneous sparse approximation has been extensively studied and many interesting and elegant results were obtained under different performance metrics. Among them, the result most related to our work is [19, Theorem 5.3], which presents a sufficient condition for simultaneous sparse pattern recovery. The difference between our work and [19], [20] lies in two aspects. First, the problem considered in this paper is more general than that of [19], [20] since simultaneous sparse approximation is a special form of block-sparse signal recovery with the measurement matrix having a block-diagonal structure and identical diagonal blocks. Second, block-sparsity is exploited in our paper to improve the recovery ability of dealing with a higher sparsity level, whereas for [19], [20], the simultaneous sparse approximation does not lead to a more relaxed condition on the dictionary as compared with the conventional single vector sparse approximation.

## V. NUMERICAL RESULTS

We present numerical results to illustrate the sparsity pattern recovery performance of the BOMP algorithm. In the simulations, the dictionary is randomly generated with each entry independently drawn from Gaussian distribution with zero mean and unit variance. We then normalize each column of the dictionary to satisfy the unit-norm constraint. The dictionary is divided into consecutive blocks of length  $d$ . The support set of the block-sparse signal is randomly chosen according to a uniform distribution, and the signals on the support set are i.i.d. Gaussian random variables with zero mean and unit variance. The measurement noise vector is randomly generated with each entry drawn from Gaussian distribution with zero mean and variance  $\sigma_w^2$ .

To show the effectiveness of the BOMP algorithm, we compare it with the OMP algorithm that does not take block-sparsity into account. Fig. 1 shows the sparsity pattern recovery success rate as a function of the block-sparsity level,  $K$ . The sparsity pattern recovery is considered successful only if the algorithm determines all the correct support indices in the first  $K$  steps for the BOMP or in the first  $Kd$  steps for the OMP, supposing the block-sparsity level,  $K$ , is known *a priori*. The results are averaged over 1000 Monte Carlo runs, with the dictionary, the signal, and the noise randomly generated for each run. From Fig. 1, we observe that for both the BOMP and the OMP algorithms, the success rate decreases as the block-sparsity level,  $K$ , increases. Also, it can be seen

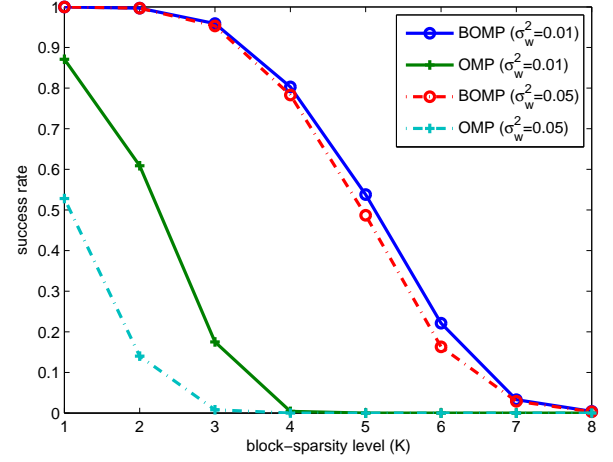


Fig. 1. Sparsity pattern recovery success rates of OMP and BOMP algorithms vs. block sparsity level,  $m = 40$ ,  $n = 400$ ,  $d = 4$ , and  $L = 100$ .

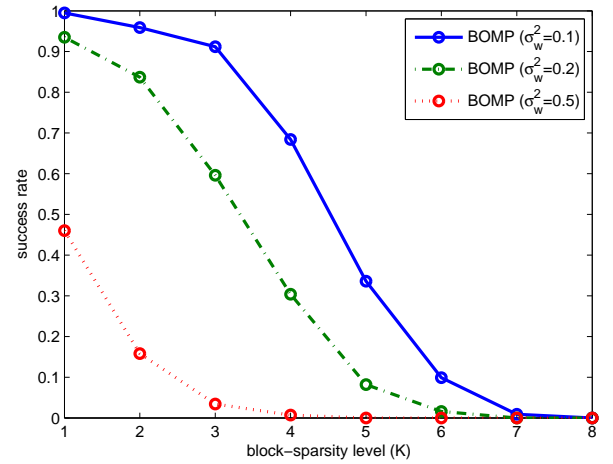


Fig. 2. Sparsity pattern recovery success rate of BOMP algorithm vs. block sparsity level,  $m = 40$ ,  $n = 400$ ,  $d = 4$ , and  $L = 100$ .

that the BOMP algorithm presents a significant performance improvement over the OMP. The result corroborate our theoretical claim that exploiting block-sparsity can lead to an improved recovery ability. Fig. 2 depicts the success rate of the BOMP algorithm under different noise power levels. We see that as the noise power increases, the recovery performance degrades. This observation is quite intuitive and coincides with our theoretical result since a higher noise power calls for a stricter requirement on the measurement matrix in order to satisfy the condition (22).

## VI. CONCLUSION

We studied the problem of recovering the sparsity pattern of block-sparse signals from noise-corrupted measurements. Our results showed that even in the presence of noise, the block-sparsity pattern can still be completely recovered via a block-version of the OMP algorithm when certain conditions are satisfied. Also, our analysis revealed that exploiting block-sparsity can lead to a guaranteed recovery for a potentially

higher sparsity level. This theoretical claim was also corroborated by our numerical results.

## APPENDIX A PROOF OF THEOREM 1

To prove Theorem 1, we only need to prove that (19) holds for  $0 \leq k \leq K-1$  given the condition (22) satisfied. To this goal, we first derive an upper bound on the second term on the left-hand side (L.H.S.) of (19).

The numerator of the second term on the L.H.S. of (19) is upper bounded by

$$\|\mathbf{A}_z^T \tilde{\mathbf{w}}\|_{2,\infty} \leq \|\mathbf{A}^T \tilde{\mathbf{w}}\|_{2,\infty} = \omega \quad (28)$$

To derive an upper bound on the second term on the L.H.S. of (19), we need to obtain a lower bound on its denominator in terms of the block coherence parameter  $\mu_B$  and the sub-coherence parameter  $\nu$ . We have

$$\begin{aligned} \|\mathbf{A}_{nz}^T \tilde{\mathbf{r}}_k\|_{2,\infty} &= \|\mathbf{A}_{nz}^T (\Phi_2 \phi_2 - \mathcal{P}_{\Phi_1} \Phi_2 \phi_2)\|_{2,\infty} \\ &\stackrel{(a)}{\geq} \|\mathbf{A}_{nz}^T \Phi_2 \phi_2\|_{2,\infty} - \|\mathbf{A}_{nz}^T \mathcal{P}_{\Phi_1} \Phi_2 \phi_2\|_{2,\infty} \end{aligned} \quad (29)$$

where (a) comes from the general mixed  $\ell_2/\ell_p$ -norm triangle inequality. The first term on the right-hand side (R.H.S.) of (29) can be further lower bounded as

$$\begin{aligned} &\|\mathbf{A}_{nz}^T \Phi_2 \phi_2\|_{2,\infty} \\ &= \max_{i \in I_1} \|\mathbf{A}_i^T \Phi_2 \phi_2\|_2 = \max_{i \in I_1} \left\| \sum_{j=k+1}^K \mathbf{A}_i^T \mathbf{A}_{l_j} \tilde{\mathbf{x}}_{l_j} \right\|_2 \\ &\geq \max_{i \in \{l_{k+1}, \dots, l_K\}} \left\{ \|\mathbf{A}_i^T \mathbf{A}_i \tilde{\mathbf{x}}_i\|_2 - \sum_{\{j|l_j \neq i, k+1 \leq j \leq K\}} \|\mathbf{A}_i^T \mathbf{A}_{l_j} \tilde{\mathbf{x}}_{l_j}\|_2 \right\} \\ &\stackrel{(a)}{\geq} \max_{i \in \{l_{k+1}, \dots, l_K\}} \left\{ (1 - (d-1)\nu) \|\tilde{\mathbf{x}}_i\|_2 - d\mu_B \sum_{\{j|l_j \neq i, k+1 \leq j \leq K\}} \|\tilde{\mathbf{x}}_{l_j}\|_2 \right\} \\ &\geq (1 - (d-1)\nu - (K-k-1)d\mu_B) \max_{i \in \{l_{k+1}, \dots, l_K\}} \|\tilde{\mathbf{x}}_i\|_2 \end{aligned} \quad (30)$$

where (a) comes from the fact that  $\lambda_{\min}(\mathbf{A}_i^T \mathbf{A}_i) \geq 1 - (d-1)\nu$  (this fact comes directly from the Gershgorin Circle Theorem), and  $\rho(\mathbf{A}_i^T \mathbf{A}_j) \leq d\mu_B$  for  $i \neq j$ . On the other hand, the second term on the R.H.S. of (29) can be upper bounded by (Please see Appendix B for the detailed derivation)

$$\|\mathbf{A}_{nz}^T \mathcal{P}_{\Phi_1} \Phi_2 \phi_2\|_{2,\infty} \leq d\mu_B (K-k) \max_{i \in \{l_{k+1}, \dots, l_K\}} \|\tilde{\mathbf{x}}_i\|_2 \quad (31)$$

Combining (29)–(31), (29) is further lower bounded by

$$\begin{aligned} &\|\mathbf{A}_{nz}^T \tilde{\mathbf{r}}_k\|_{2,\infty} \\ &\geq (1 - (d-1)\nu - (2K-2k-1)d\mu_B) \max_{i \in \{l_{k+1}, \dots, l_K\}} \|\tilde{\mathbf{x}}_i\|_2 \\ &\geq (1 - (d-1)\nu - (2K-2k-1)d\mu_B) x_{b,\min} \end{aligned} \quad (32)$$

Since (18) is a necessary condition for (19), we should always have  $1 - (d-1)\nu - (2K-1)d\mu_B > 0$ . Therefore we can

guarantee that the above derived lower bound is positive. Consequently an upper bound on the second term on the L.H.S. of (19) can be derived and given as

$$\begin{aligned} \frac{\|\mathbf{A}_z^T \tilde{\mathbf{w}}\|_{2,\infty}}{\|\mathbf{A}_{nz}^T \tilde{\mathbf{r}}_k\|_{2,\infty}} &\leq \frac{\omega}{(1 - (d-1)\nu - (2K-2k-1)d\mu_B) x_{b,\min}} \\ &\leq \frac{\omega}{(1 - (d-1)\nu - (2K-1)d\mu_B) x_{b,\min}} \end{aligned} \quad (33)$$

We see that the first and the second term on the L.H.S. of (19) are respectively upper bounded by (17) and (33). Therefore (19) is guaranteed if the summation of these two upper bounds are smaller than unity, i.e.

$$\begin{aligned} &\frac{Kd\mu_B}{1 - (d-1)\nu - (K-1)d\mu_B} \\ &+ \frac{\omega}{(1 - (d-1)\nu - (2K-1)d\mu_B) x_{b,\min}} < 1 \end{aligned} \quad (34)$$

A further transformation easily shows that (34) and (22) are equivalent (note that the condition  $1 - (d-1)\nu - (2K-1)d\mu_B > 0$  has to be explicitly indicated to assure (18) and to assure the positiveness of the lower bound (32)). The proof is completed here.

## APPENDIX B DERIVATION OF EQUATION (31)

Clearly we have

$$\|\mathbf{A}_{nz}^T \mathcal{P}_{\Phi_1} \Phi_2 \phi_2\|_{2,\infty} = \max_{i \in I_1} \|\mathbf{A}_i^T \mathcal{P}_{\Phi_1} \Phi_2 \phi_2\|_2 \quad (35)$$

We consider two different cases. If  $\mathbf{A}_i$  is a column-block of  $\Phi_1$ , i.e.  $i \in \{l_1, \dots, l_k\}$ , then for any index  $i$ , we have

$$\begin{aligned} \|\mathbf{A}_i^T \mathcal{P}_{\Phi_1} \Phi_2 \phi_2\|_2 &\stackrel{(a)}{=} \|\mathbf{A}_i^T \Phi_2 \phi_2\|_2 = \left\| \sum_{j=k+1}^K \mathbf{A}_i^T \mathbf{A}_{l_j} \tilde{\mathbf{x}}_{l_j} \right\|_2 \\ &\leq \sum_{j=k+1}^K \|\mathbf{A}_i^T \mathbf{A}_{l_j} \tilde{\mathbf{x}}_{l_j}\|_2 \leq d\mu_B \sum_{j=k+1}^K \|\tilde{\mathbf{x}}_{l_j}\|_2 \\ &\leq d\mu_B (K-k) \max_{i \in \{l_{k+1}, \dots, l_K\}} \|\tilde{\mathbf{x}}_i\|_2 \end{aligned} \quad (36)$$

where (a) comes from the fact that  $\Phi_1^T \mathcal{P}_{\Phi_1} = \Phi_1^T \Phi_1 (\Phi_1^T \Phi_1)^{-1} \Phi_1^T = \Phi_1^T$ , and therefore  $\mathbf{A}_i^T \mathcal{P}_{\Phi_1} = \mathbf{A}_i^T$  for  $i \in \{l_1, \dots, l_k\}$ . On the other hand, if  $\mathbf{A}_i$  is a column-block of  $\Phi_2$ , i.e.  $i \in \{l_{k+1}, \dots, l_K\}$ . We show that

$$\max_{i \in \{l_1, \dots, l_k\}} \|\mathbf{A}_i^T \mathcal{P}_{\Phi_1} \Phi_2 \phi_2\|_2 \geq \max_{i \in \{l_{k+1}, \dots, l_K\}} \|\mathbf{A}_i^T \mathcal{P}_{\Phi_1} \Phi_2 \phi_2\|_2 \quad (37)$$

To this goal, let  $\mathbf{z} \triangleq \Phi_1^\dagger \Phi_2 \phi_2 = [\mathbf{z}_1^T \dots \mathbf{z}_k^T]^T$ , the term on the L.H.S. of (37) is lower bounded as

$$\begin{aligned}
& \max_{i \in \{l_1, \dots, l_k\}} \|\mathbf{A}_i^T \mathcal{P}_{\Phi_1} \Phi_2 \phi_2\|_2 = \max_{i \in \{l_1, \dots, l_k\}} \|\mathbf{A}_i^T \Phi_1 \mathbf{z}\|_2 \\
& = \max_{i \in \{l_1, \dots, l_k\}} \left\| \sum_{j=1}^k \mathbf{A}_i^T \mathbf{A}_{l_j} \mathbf{z}_j \right\|_2 \\
& \stackrel{(a)}{\geq} \|\mathbf{A}_{l_q}^T \mathbf{A}_{l_q} \mathbf{z}_q\|_2 - \sum_{\{j|j \neq q, 1 \leq j \leq k\}} \|\mathbf{A}_{l_q}^T \mathbf{A}_{l_j} \mathbf{z}_j\|_2 \\
& \geq (1 - (d-1)\nu) \|\mathbf{z}_q\|_2 - d\mu_B \sum_{\{j|j \neq q, 1 \leq j \leq k\}} \|\mathbf{z}_j\|_2 \\
& \geq (1 - (d-1)\nu - (k-1)d\mu_B) \|\mathbf{z}_q\|_2 \quad (38)
\end{aligned}$$

where in (a), the index  $q$  is chosen such that  $\mathbf{z}_q$  has the maximum  $\ell_2$ -norm among  $\{\mathbf{z}_i\}_{i=1}^k$ . The term on the R.H.S. of (37) is upper bounded by

$$\begin{aligned}
& \max_{i \in \{l_{k+1}, \dots, l_K\}} \|\mathbf{A}_i^T \mathcal{P}_{\Phi_1} \Phi_2 \phi_2\|_2 = \max_{i \in \{l_{k+1}, \dots, l_K\}} \|\mathbf{A}_i^T \Phi_1 \mathbf{z}\|_2 \\
& = \max_{i \in \{l_{k+1}, \dots, l_K\}} \left\| \sum_{j=1}^k \mathbf{A}_i^T \mathbf{A}_{l_j} \mathbf{z}_j \right\|_2 \leq \max_{i \in \{l_{k+1}, \dots, l_K\}} \sum_{j=1}^k \|\mathbf{A}_i^T \mathbf{A}_{l_j} \mathbf{z}_j\|_2 \\
& \leq d\mu_B \sum_{j=1}^k \|\mathbf{z}_j\|_2 \leq kd\mu_B \|\mathbf{z}_q\|_2 \quad (39)
\end{aligned}$$

Since we have  $1 - (d-1)\nu - (2K-1)d\mu_B > 0$  in order to assure the condition (18) to be satisfied, we can easily verify that the following always holds for  $0 \leq k < K$

$$(1 - (d-1)\nu - (k-1)d\mu_B) > kd\mu_B \quad (40)$$

The inequality (37) comes directly by combining (38–40). Therefore the second term on the R.H.S. of (29) is upper bounded by

$$\begin{aligned}
\|\mathbf{A}_{\text{nz}}^T \mathcal{P}_{\Phi_1} \Phi_2 \phi_2\|_{2,\infty} &= \max_{i \in I_1} \|\mathbf{A}_i^T \mathcal{P}_{\Phi_1} \Phi_2 \phi_2\|_2 \\
&= \max_{i \in \{l_1, \dots, l_k\}} \|\mathbf{A}_i^T \mathcal{P}_{\Phi_1} \Phi_2 \phi_2\|_2 \\
&\leq d\mu_B (K - k) \max_{i \in \{l_{k+1}, \dots, l_K\}} \|\tilde{\mathbf{x}}_i\|_2 \quad (41)
\end{aligned}$$

where the last inequality comes from (36).

## REFERENCES

- [1] A. J. Miller, *Subset selection in regression*. New York: Chapman&Hall, 1990.
- [2] N. Meinshausen and P. Bühlmann, “High-dimensional graphs and variable selection with the Lasso,” *Ann. Statist.*, pp. 1436–1462, 2006.
- [3] E. Candès and T. Tao, “Decoding by linear programming,” *IEEE Trans. Information Theory*, no. 12, pp. 4203–4215, Dec. 2005.
- [4] J. J. Fuchs, “Recovery of exact sparse representations in the presence of bounded noise,” *IEEE Trans. Information Theory*, vol. 51, no. 10, pp. 3601–3608, Oct. 2005.
- [5] J. A. Tropp, “Just relax: convex programming methods for identifying sparse signals in noise,” *IEEE Trans. Information Theory*, vol. 52, no. 3, pp. 1030–1051, Mar. 2006.
- [6] M. J. Wainwright, “Sharp thresholds for high-dimensional and noisy sparsity recovery using  $\ell_1$ -constrained quadratic programming (lasso),” *IEEE Trans. Information Theory*, vol. 55, no. 5, pp. 2183–2202, May 2009.
- [7] —, “Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting,” *IEEE Trans. Information Theory*, vol. 55, no. 12, pp. 5728–5741, Dec. 2009.

- [8] A. K. Fletcher, S. Rangan, and V. K. Goyal, “Necessary and sufficient conditions for sparsity pattern recovery,” *IEEE Trans. Information Theory*, vol. 55, no. 12, pp. 5758–5772, Dec. 2009.
- [9] M. Mishali and Y. C. Eldar, “Blind multi-band signal reconstruction: compressed sensing for analog signals,” *IEEE Trans. Signal Processing*, vol. 57, no. 3, pp. 993–1009, Mar. 2009.
- [10] R. Gribonval and E. Bacry, “Harmonic decomposition of audio signals with matching pursuit,” *IEEE Trans. Signal Processing*, vol. 51, no. 1, pp. 101–111, Jan. 2003.
- [11] Y. C. Eldar, P. Kuppinger, and H. Bölcskei, “Block-sparse signals: uncertainty relations and efficient recovery,” *IEEE Trans. Information Theory*, vol. 58, no. 6, pp. 3042–3054, June 2010.
- [12] Y. C. Eldar and M. Mishali, “Robust recovery of signals from a structured union of subspaces,” *IEEE Trans. Information Theory*, vol. 55, no. 11, pp. 5302–5316, Nov. 2009.
- [13] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, “Model-based compressive sensing,” *IEEE Trans. Information Theory*, vol. 56, no. 4, pp. 1982–2001, Apr. 2010.
- [14] S. Chen, S. A. Billings, and W. Luo, “Orthogonal least squares methods and their application to non-linear system identification,” *Int. J. Control*, vol. 50, no. 5, pp. 1873–1896, Nov. 1989.
- [15] Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad, “Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition,” in *Proc. 27th Annu. Asilomar Conf. Signals, Systems, and Computers*, vol. 1, Pacific Grove, CA, Nov. 1993, pp. 40–44.
- [16] J. A. Tropp, “Greed is good: algorithmic results for sparse approximation,” *IEEE Trans. Information Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.
- [17] J. A. Tropp and A. C. Gilbert, “Signal recovery from random measurements via orthogonal matching pursuit,” *IEEE Trans. Information Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.
- [18] A. K. Fletcher and S. Rangan, “Orthogonal matching pursuit from noisy measurements: a new analysis,” in *Twenty-Third Annual Conference on Neural Information Processing Systems*, Vancouver, B.C., Canada, 2009.
- [19] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, “Algorithms for simultaneous sparse approximation part I: greedy pursuit,” *Signal Processing*, vol. 86, no. 3, pp. 572–588, 2006.
- [20] R. Gribonval, H. Rauhut, K. Schnass, and P. Vandergheynst, “Atoms of all channels, unite! average case analysis of multi-channel sparse recovery using greedy algorithms,” *The Journal of Fourier Analysis and Applications*, vol. 14, no. 5, 2008.